Trading Strategy for Bitcoin and Gold

Summary

As the world shifts its focus to addressing real-world problems using computer programming, creative ways of problem solving have started to develop rapidly. This includes the implementation of algorithmic trading in financial investments such as stocks, cryptocurrency, and other markets. However, because such assets are volatile, it is challenging, even for experts, to predict their prices. For this, we built a model to achieve the goal of making more accurate price predictions and profitable market orders (buy/sell/hold).

In this paper, an auto-regressive integrated moving average (ARIMA) model, linear regression, and polynomial regression are used to make price predictions. From the data given, each model/regression uses a different method to predict prices. At the end, each method provides a slightly different price curve. However, due to the slight differences in price predictions, they lead to significant differences in total return after implementing the trading indicators.

The trading indicators used in this paper are the standard deviation channels and exponential ratios. The size of the standard deviation channels is calculated from the standard deviations from each price prediction curve (polynomial, linear, and ARIMA). Additionally, the standard deviation channels are set to have seven-day windows, meaning anytime a price point goes below or above the channel in that seven-day period, it triggers a trading action to buy or sell. Once a trading action is triggered, to determine the proportions of our holdings to sell or of our cash to buy, we use the exponential ratios. For instance, when the exponential ratio is higher, we buy/sell more.

Given the $1,000 initial investment and day-to-day pricing data for gold and bitcoin, a good model is said to beat the market by consistently making profitable market orders. This is done by selling at higher price points and buying at lower price points while considering the 1% and 2% transactions costs for gold and bitcoin. Based on this goal, our model finds the total return to be around $610,000, which is close to nine times higher than the expected market return ($∼72,000) if no actions are made from our initial $1,000 investment. This result was generated by using the ARIMA model as the price-predicting strategy, and a combination of standard deviation channel and exponential ratio is used as the market indicators. Thus, this also proves that the ARIMA model is the optimal price-predicting strategy compared to linear and polynomial regression.

Contents

1 Introduction

1.1 Problem Background

Algorithmic trading is a technical method used to make trading decisions based on predetermined trading guidelines regarding prices, time, and quantity [5]. Therefore, a typical model that aims to maximize total return through a certain amount of time requires two main steps. First, a method should be developed to predict the future prices based on prices in the previous days. Second, based on the price predictions, another method should be used to make market decisions to buy, sell, or hold.

However, because there is a significant number of factors and indicators that affect the prices and traders' decisions, it requires researchers to find a combination of models that best fit their circumstances. Hence, we decided to use auto-regressive integrated moving average (ARIMA), linear regression, and polynomial regression to predict the prices and standard deviation channels to act upon such prices for maximum total return.

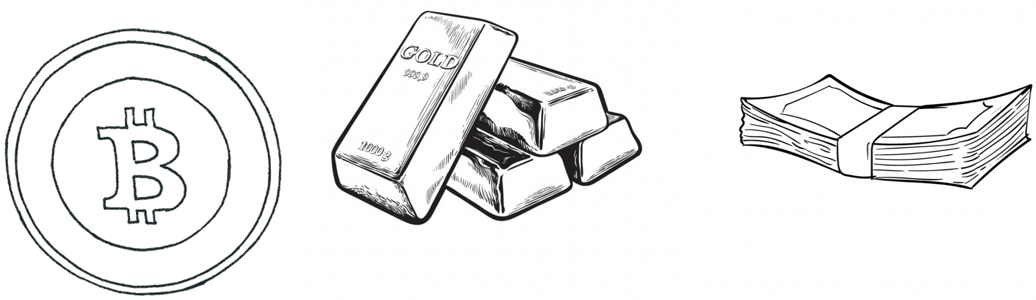

Figure 1: Hand Drawings of Bitcoin, Gold, and Cash

1.2 Problem Statement

No two market-traded assets follow the exact price adjustments or fluctuations. As a result, investors will attempt to place their funds in a market that will give them an optimal return on investment. By utilizing automated intelligence and advanced trading indicators to assess market price action, investors can maximize their profit. In acknowledgement of the given conditions, the following issues are required to be discussed in the paper:

- An accurate price-predicting strategy
- A mathematical model identifying the best buying and selling prices based off of market indicators
- An optimal trading strategy relative to transaction cost
- Capital dispersed between cash, gold, and bitcoin simultaneously based on the best setups of the individual markets

2 Assumptions and Justifications

The model is built on the following assumptions to simplify problems:

1. **Trading will be executed with no influence from outside factors.** This is to eliminate the impact on the trader's decision based on social factors (financial crisis), personal reasons (financial situation and personal preferences), and other possible economic and political causes.

2. **Trades without sufficient funds/capital cannot be executed.** This is to ensure the trader is not allowed to be in debt while trading these assets.

3. **Fractional shares of bitcoin and gold may be purchased.** To be more realistic with the current trading system, fractional shares are allowed. This also benefits the current situation in the problem since $1000 is not a big amount of initial investment considering the starting prices of gold and bitcoin are $1324.6 and $621.65.

    - **In addition with this assumption, minimum transaction amount for gold must be greater than $1 and $0.50 for bitcoin.** This is to avoid having partial cents due to the 1% and 2% commission fees.

4. **Arbitrage trading is not allowed.** Arbitrage is the strategy that an investor buys an asset from one market with a lower price and sells it in another market for a higher price. Not allowing this is to ensure the investor uses only the up-to-date pricing data given from the problem.

3 Literature Review

Due to a magnificent number of algorithmic trading strategies, this section examines a few useful methods for making price predictions and market orders based on indicators. Specifically, machine learning (ML), deep neural network (DNN), auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA) are different price prediction methods we previously considered. For market indicators, we studied linear regression model, polynomial regression model, fibonacci retracement, moving average convergence divergence, and relative strength index.

3.1 Price Prediction

Lv et al.'s research compares six traditional machine learning (ML) algorithms with six deep neural network (DNN). For price predictions, they used forty-four indicators to determine market orders (buy/sell/hold) [4]. ML is often used to predict prices by training the model to identify pricing trends. The six ML algorithms used in Lv et al.'s work were support vector machine (SVM), random forest (RF), logistic regression (LR), naïve Bayes model (NB), classification and regression tree (CART), and Extreme Gradient Boosting algorithm (XGB) [4]. Though ML has been proven to be effective in price predictions, it has shown that, when compared to the trading performance using DNN algorithms, ML algorithms are more sensitive to transaction costs. On the other hand, DNN is an extension of ML since it utilizes hidden layers that evaluate the significance of the input variables in the original ML algorithm. Because of this, DNN considers transaction costs with a greater focus, thus leading to a better trading performance [4].

Though ML and DNN are currently two of the most popular methods for making price predictions, they do not fit the goal and data of this project. This is due to the fact that they have to utilize the entire data set. However, because we are only allowed to use the pricing data up to each day, using such the data from outside of the time range is prohibited. Thus, Auto Regressive Moving Average (ARMA) and auto regressive integrated moving average (ARIMA) will be closely studied to determine the most suitable price prediction model for this project.

Both ARIMA and ARMA are time-series models that forecast the target variable based on current data. Since they do not require training using the entire data set like ML and DNN do, ARIMA and ARMA are more suitable for the purpose of this project. More specifically, ARMA is a stationary schotastic process whereas ARIMA is non-stationary [7]. ARIMA is also a generalization of ARMA. Additionally, in Valipour et al.'s research, ARIMA was found to have a better performance than ARMA because of the non-stationary time series in the forecasting phase. We will also utilize linear and polynomial regression techniques for price forecasting.

3.2 Market Indicators

With accurate price predictions, traders can then act upon such predictions based on market indicators that would maximize their total return. Thus, because each indicator has a specific purpose, we will study different indicators to better understand which ones to implement in our model.

Regression techniques are extensively used in economics to study the relationship between multiple variables. Specifically the tool of polynomial regression and its ability for creating a flexible curve is effective for analyzing the volatile cryptocurrency and gold markets. Polynomial regression allows for a strong analysis of trend direction, momentum, and establishment for areas of support/resistance. Linear regression may also be used for many of the same applications but in a linear fashion. Utilizing standard deviation channels along with regression techniques increases the confidence interval based on price actions. This tells from a quantitative perspective of how likely the price is to return to the mean (the regression line) [6].

Other indicators such as relative strength index (RSI), moving average convergence divergence (MACD), and Fibonacci retracement based on the golden ratio have been considered. Drawbacks of these indicators are as follows. RSI is typically best paired with many other indicators, and it is highly unreliable when used by itself due to its emphasis on moving averages. The MACD indicator has many strengths such as spotting trend reversals, but fails to perform in sideways markets. Because the MACD indicator is reliant on underlying price points, a moving average volume oscillator will give a better estimation of target signals. Fibonacci retracement was also considered, but due to its dependence on price noise, it is found to cause errors and false signals as it is not a trend indicator. The Fibonacci retracement is of more use as a lagging indicator which can establish support and resistance lines [1, 2, 3].

4 Model Building

4.1 Method

For the methods and papers reviewed in the Literature Review section, we conclude that using an ARIMA model and linear/polynomial regression combined with market indicators (standard deviation channels and exponential ratios) works best for the purpose of this project. The ARIMA model is chosen because our data is non-stationary. This is shown by the changing statistical properties (e.g. mean, standard deviation, autocorrelation). Polynomial and linear regression are chosen for their effectiveness in analyzing the volatile cryptocurrency and gold markets. As one of the market indicators, standard deviation channels are used because, along with linear/polynomial regression, they can increase the confidence interval while giving the trader information of when the trader should trade. Lastly, exponential ratios show how much of the trader's asset should be traded, which will be discussed in a later subsection.

In addition to the three models mentioned, we create a baseline model with limited predictability to benchmark the performance of our selected models.

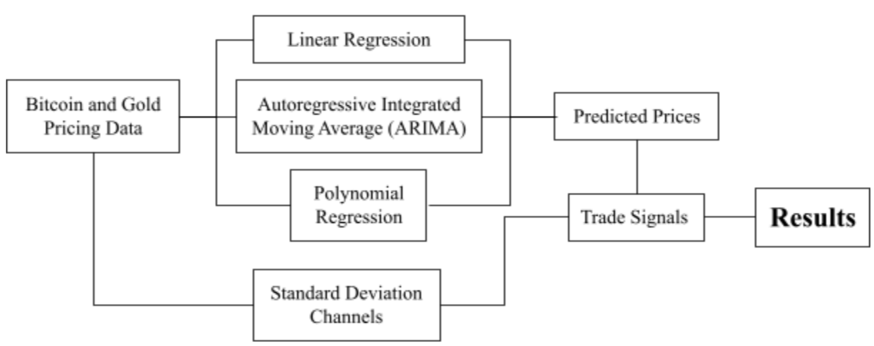

Figure 2: Model Building Process

4.2 Linear Regression, Polynomial Regression, and Standard Deviation Channels

While linear regression is a classical method to obtain the rate of change of a variable, polynomial regression analysis involves identifying the relationship between a dependent variable and other independent variables. The regression line consists of successive power terms, the highest order term plus all lower order terms. This equation takes the form of

$$f(x) = c_0 + c_1 x + c_2 x^2 + ... + c_n x^n \tag{1}$$

where n is the degree of the polynomial and c is a set of coefficients.

Standard deviation channels can be formed from both the linear and polynomial regression lines. The sum of least-squares equation creates the line of best fit through all the data points and parameter estimates to minimize variance from the actual data points (Equation 2). To form standard deviation channels, lines are plotted based on the line of best fit from the sum of least-squares equation. When the lines are plotted one standard deviation above and below the trend line, we would expect 68% of the interval data points to fall within the range of the channel. If the number of standard deviations is increased to two for the upper and lower boundaries, 95% of the interval data points are expected to fall within the channel (Figure 3). For the baseline model, the channel boundaries are defined as two standard deviations away from last day's price, which does not involve any prediction. Such design facilitates the evaluation and comparison process for the other models.

For linear and polynomial regressions, the channel boundaries are obtained by adding or subtracting two standard deviations from the predicted price using linear or polynomial regression. Therefore, trading actions may be triggered when the price point breaks one of these channel boundaries, where the price is likely to move back inside of one of the channels unless a major trend reversal is taking place (Figure 4). Lastly, a time-window of seven days and thirty-five days for the standard deviation channels are both tested and analyzed. As a result, we found that the thirty-five-day standard deviation channel gives us more accurate long-term trade signals while the seven-day channel gives us more accurate short-term trade signals.

$$\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

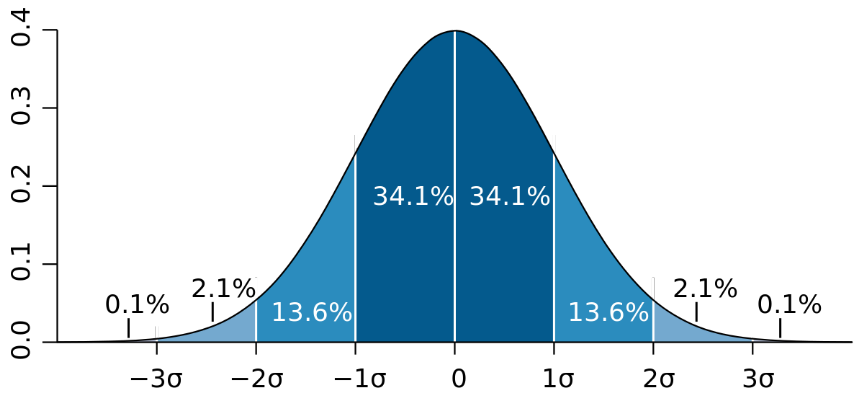

Figure 3: Standard Deviation and Confidence Level

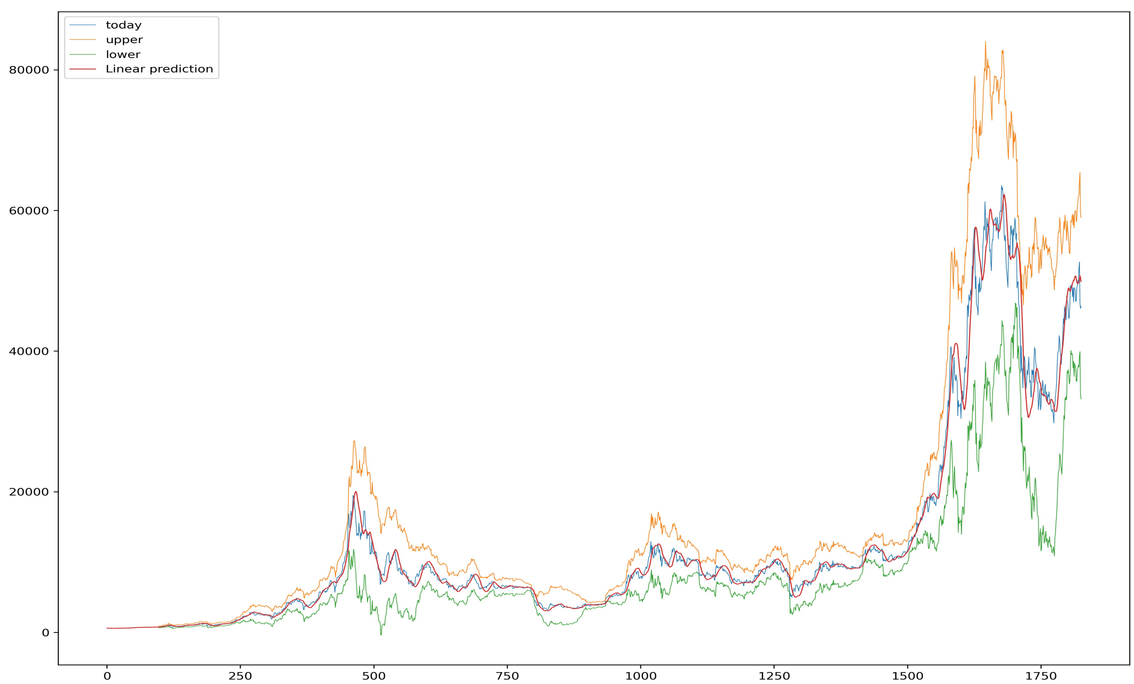

Figure 4: Standard Deviation Channel

4.3 ARIMA

As previously mentioned, the auto-regressive integrated moving average (ARIMA) is chosen due to its accuracy for non-stationary data [8]. Essentially, ARIMA's equation is expressed as:

$$ARIMA(p,d,q) = y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3}$$

where:

- p is the lag order of the auto-regressive process
- d is the order of differencing
- q is the order of the moving average process
- t is time
- c is a constant
- ϕ and θ are weights of the lagged observations
- ϵ is noise

In order to apply this model, we should show that the given data has trends, changes in variance, and are non-stationary. As shown, the autocorrelation of the price of bitcoin hardly moves away from 1.0 and stays outside the shaded area, indicating that the data is non-stationary.

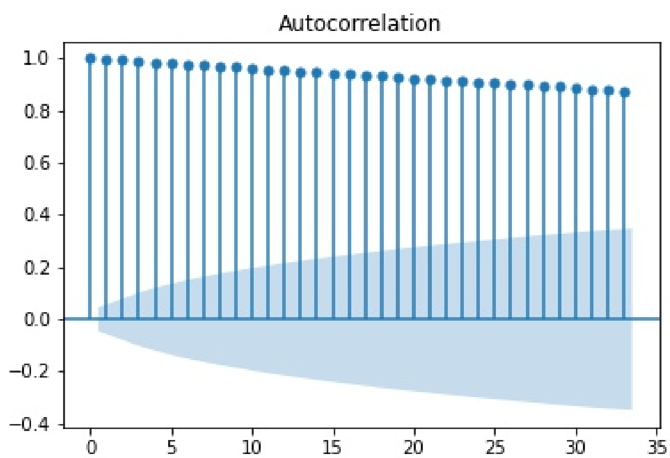

Figure 5: Autocorrelation for Price of Bitcoin

Since we are not provided nor allowed to utilize other data resources to train a prediction model for the price of either currencies, we learn from Wirawan et al. that an ARIMA(4, 1, 4) model would yield the highest accuracy of predicting bitcoin and gold prices [8]. Similar to linear and polynomial regression, we use the past data within our time-window as input for ARIMA(4, 1, 4) to predict each day's price and apply the standard deviation channels to determine transactions.

4.4 Proportions and Commission Fees

Proportion of cash spent and the ratio of cash spent on gold versus bitcoin are essential factors to consider when transacting. In addition, commission fees, which functions as the cost of each transaction also impacts the profit return. In this section, we will discuss how we take these factors into careful consideration.

4.4.1 Trading Ratio

The trading ratio determines how much cash to buy and how much shares to sell when selling. We develop three methods to determine the trading ratio based on how many standard deviations, denoted by $|z|$, is today's price from the predicted price obtained by the different prediction models. First, we fix the ratio to be 0.5 for all transactions when the price is outside the channel. Second, we use a piece-wise (linear) equation to calculate the ratio, which is displayed below.

$$trading\,ratio = \begin{cases} 0.25 & 2 \leq |z| \leq 3 \\ 0.5 & 3 \leq |z| \leq 4 \\ 0.75 & |z| \geq 5 \\ 1 & otherwise \end{cases}$$

The third approach is to apply a non-linear relationship between $|z|$ and trading ratio when $|z|$ is greater than 2.

$$1 - \exp\left(\frac{1-|z|}{\lambda}\right), 0 < \lambda < 1 \tag{4}$$

The trading ratio increases more drastically as λ decreases, which functions as a risk factor. If one is confident about the predictability of model, λ could be set smaller.

4.4.2 Gold-Bitcoin Investment Ratio

When determining investment strategies amongst bitcoin, gold, and cash, we will utilize the slope of the respective trend lines when assets are plotted against the US dollar. If the trend for an asset has a strong setup, a positive slope, and the other a weak slope, in the short term all funds will be allocated with a majority to the stronger setup. If the short term trends for both Bitcoin and gold are in the negative direction, loses will be minimized with a transfer of funds to cash.

The investment ratio between gold and bitcoin is determined by the trends of the price of both currencies. If both currencies are enjoying a considerable ascent, we will compare the growth rate between them, given by

$$Growth\ Rate = \frac{Price_t - mean_{window}(Price)}{mean_{window}(Price)}$$

The investment ratio is thus given by

$$Investment\ Ratio = \frac{Growth\ Rate(Gold)}{Growth\ Rate(Bitcoin)}$$

4.4.3 Commission Fees

The commission fees $\alpha_{gold} = 1\%$ and $\alpha_{bitcoin} = 2\%$ are considered the cost of each transaction. For a transaction to be considered profitable, in addition to having standard deviations greater than 2, the profit has to be greater than the cost given below.

$$Profit > Price * Shares * \alpha$$

5 Model Validating

In this section, we present the profit returns of our four prediction models across three different methods to calculate trading ratios. We also observe and analyze the relationship between MAPE (Mean Absolute Percentage of Error), which evaluates the predictability, and total return.

5.1 Results

As mentioned in Section 4, we apply standard deviation channels on four prediction methods: a baseline model without prediction and three prediction models using linear regression, polynomial regression and ARIMA(4, 1, 4). For each model, we test the performance with fixed, linear and exponential trading ratios. The result of this ablation study is displayed in Table 1. The market price is included for comparison.

The baseline model fails to exceed market price of bitcoin on a fixed trading ratio, but successfully surpasses the market price with linear and exponential ratio. Among the four models, **ARIMA** outperforms the other three models with all various trading ratio designs, and rewarding us a total profit of $609,099,83 with exponential ratio. The plot of daily values of assets for various methods is included.

	Market	Baseline	Lin Reg	Poly Reg	ARIMA
MAPE	X	8.07	4.73	1.95	1.34
Fixed Ratio	72,403.13	57,028.51	96,754.21	160,021.67	185,994.17
Linear-Growth Ratio	72,403.13	75,890.59	182,122.35	337,527.21	466,294.32
Exponential Ratio	**72,403.13**	**106,553.17**	**270,851.13**	**429,818.31**	**609,099.83**

Table 1: Results of Ablation Study Across Trading Ratio Design and Models

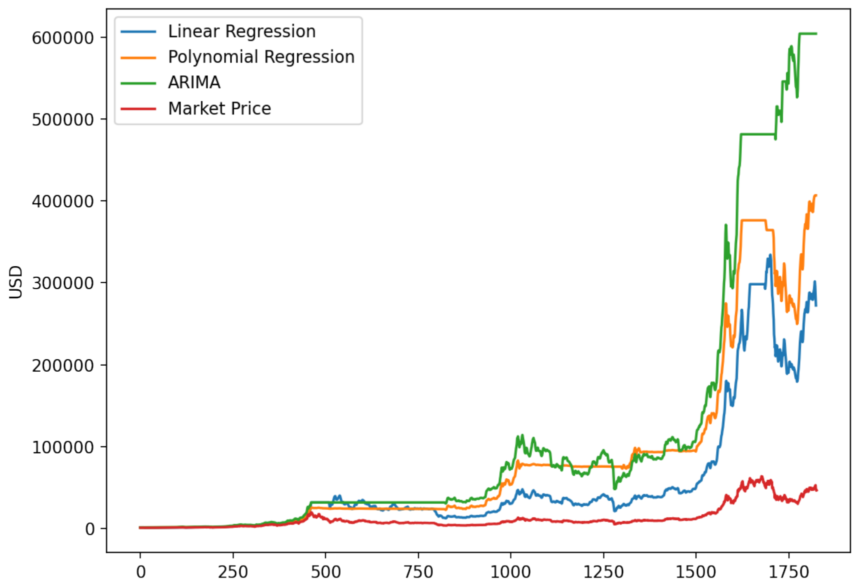

Figure 6: Result Comparison

5.2 Analysis

ARIMA with exponential ratio Interestingly, all models yield the most favorable total return with exponential ratio and outperform other trading ratio designs significantly. Exponential ratio (Equation 2) grows more drastically than linear, meaning that we would be more aggressive when we observe a turn in the market.

The mean absolute percentage of error also demonstrates strong correlation between predictability and total return. As MAPE decreases, the return increases considerably, meaning that our model's performance is strongly correlated with how well it can predict the data. In another word, data that is difficult to predict could probably impact the total return.

6 Strengths and Weaknesses

6.1 Strengths

1. **Easy for implementation.** Algorithmic trading creates a hands-free environment in which indicators can perform an analysis for trend direction, momentum, and establishment for areas of support/resistance. Our model also utilizes basic methods, making it easy to understand reproduce. In addition, the model explicitly provides the trader with simple instructions to buy/sell/hold a specific proportion of an asset.

2. **Degree of flexibility of the model.** The customizable time frame for ARIMA allows us to change the sensitivity (proportions of trades) of our trading patterns based upon the volatility of the market. Since standard deviation is an extremely well-tested statistical property, it is used in a wide-range of areas. This also means that a lot of advanced market indicators, such as Bollinger Bands, are based on the fundamental idea of standard deviation. Therefore, the flexibility of this model creates an applicability of the indicator to be used across all markets.

6.2 Weaknesses

1. **Lack of analysis and testing for market indicators.** Our model only implements the regression model and standard deviation channels as indicators. However, there is a significant number of trading indicators available that our model did not consider. Implementing such indicators may improve our model, but it would also complicate our processes and interpretations of the algorithm's trading decisions. Furthermore, as mentioned before, we found two different time windows to be effective for our standard deviation channels (seven-day and thirty-five-day). One time-window of the channel was applied to both assets at one time. However, depending on the volatility of each asset, using a smaller time window on the more volatile asset and a bigger time window on the less volatile asset may generate more accurate trade signals.

2. **Lack of consideration for outside factors.** This includes sentiment analysis, which can key role in analyzing short-term trading signals, inhibits the efficiency of our model. In the age of social media, outside media factors could be adopted to greatly improve predictions for market trading.

7 Conclusion

Based on the simulated investment results, below concludes our analysis. The bitcoin and gold transaction fees, available trading hours, and investment strategies served critical to the financial return on the initial $1,000 investment. Multiple simulated investment models with detailed mathematical analysis give answers to the required questions.

Linear regression, polynomial regression, ARIMA, and standard deviation channels are investment tools used to create the maximum returns. If an initial \$1,000 investment is made in bitcoin on 9/11/2016 held until 9/10/2021, an investor would see a return of \$72,403.13. The overall 7,240% increase in investment on bitcoin and 35% for gold clearly shows bitcoin as a better investment a majority of the time. However, this is not to state that investing in gold did not occur, as gold provided a profitable return on investment when bitcoin saw heavy losses and gold had a strong outlook.

A fixed buy/sell ratio, linear-growth ratio, and exponential ratio are tested against varying indicators. In terms of profit, the exponential ratio paired with the ARIMA indicator shows maximum returns followed by polynomial regression, linear regression, and baseline in the exponential section. The linear-growth ratio proved to be second most effective with all the tested indicators followed lastly by the fixed buy/sell ratios.

References

[1] S. Bhattacharya and K. Kumar. A computational exploration of the efficacy of Fibonacci Sequences in Technical analysis and trading. *Annals of Economics and Finance*, 7:219–230, 2006.

[2] A. Bhowmick, A. Rahman, and R. M. Rahman. Performance analysis of different Recurrent Neural Network Architectures and Classical Statistical Model for Financial Forecasting: A Case Study on Dhaka Stock Exchange. 985:277–286, 2019.

[3] S. Halilbegovic, N. Čelebić, and D. Kulović. Analysis of a standalone usage and limitations of a relative strength index indicator in stock trading. *Ecoforum*, 7:1–10, 2018.

[4] D. Lv, S. Yuan, M. Li, and Y. Xiang. An empirical study of machine learning algorithms for stock daily trading strategy. *Mathematical Problems in Engineering*, 2019:1–30, 2019.

[5] M. Mathur, M. Satyam, M. Sahil, and M. Vanita. Algorithmic trading bot. *ITM Web of Conferences*, 40:1–10, 2021.

[6] E. Ostertagová. Modelling using Polynomial Regression. *Procedia Engineering*, 48:500–506, 2012.

[7] M. Valipour, M. E. Banihabib, and S. M. RezaBehbahani. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 476:433–441, 2013.

[8] M. I. Wirawan, T. Widiyaningtyas, and M. M. Hasan. Short Term Prediction on Bitcoin Price Using ARIMA Method. *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2019:260–265, 2019.

Memorandum

TO: Trader

FROM: Team 2221717

DATE: February 21st, 2022

SUBJECT: Gold and Bitcoin Trading Strategy

Introduction

Our team was tasked with creating an efficient model to determine each day if you should buy, hold, or sell your assets in the portfolio. Factors that were taken into consideration when developing this model include different transaction fees with bitcoin and gold, the restriction to only use the past stream of daily prices for future market action, and the fact that bitcoin can be traded anytime throughout the week and gold only while the market is open.

Model

To determine the best method for a market investing strategy, our team decided to utilize regression techniques, auto-regressive integrated moving average (ARIMA) model, standard deviation channels, and exponential ratios. In summary, regression techniques and the ARIMA model are used for price predicting while standard deviation channels and exponential ratios are used to decide when to trade and how much to trade for.

Regression techniques are extensively used in economics to study the relationship between multiple variables. The purpose of the regression technique serves to most accurately find the "best-fit" regression line of the actual price data for bitcoin and gold. Standard deviation channels can then be formed from these regression lines. Utilizing standard deviation channels along with regression techniques increases the confidence interval based on price actions. This tells from a quantitative perspective how likely the price is to return to the mean (the regression). Similar to regression techniques, the auto-regressive integrated moving Average (ARIMA) is used for price predictions and changes in trend variance. However, it has been proven that ARIMA offers more accurate price predictions when compared to regression techniques. Thus, the ARIMA model is chosen to be our price-predicting strategy and is combined with the two market indicators to produce maximum total return: standard deviation channel and exponential ratio.

With the standard deviation channel, once an actual price point goes below/above the standard deviation channel, a trading action is triggered, meaning we expect to be able to profit off of selling or buying at this point. Additionally, to determine how much to sell or buy, we combine the standard deviation channel with exponential ratios. In other words, when the exponential ratio is higher, meaning the actual price point is further away from the upper/lower boundary of the standard deviation channel, we would buy/sell more portions of our holdings. Vice versa, when the exponetial ratio is lower, meaning the actual price point is closer to the upper/lower boundary, we would buy/sell less. Furthermore, when both bitcoin and gold's price points are out of the standard deviation channel, meaning we should buy/sell both bitcoin and gold, we decide to buy/sell certain proportions of bitcoin or gold based on their price trends at that time. For instance, if both bitcoin and gold's price points are below the standard deviation channel, we would buy more bitcoin than gold if the bitcoin has a higher up-trend at that point. If the price stays within the channel, we suggest to not execute any market action. This is mainly due to the factor of commission fees.

Results

Given the initial \$1,000 dollar investment, the highest total return our team achieved was \$609,099.83, which is almost nine times higher than the expected market return (\$∼72,000). This is sufficient evidence to show the efficiency of our model. Besides the amount of total return, our model also proves to offer more strengths: easy for implementation, and flexible. Both of which offer significant value to you as a trader. Due to the basic methods utilized in our model, it is easy to reproduce the same algorithm to generate the same outcome. This model can also be implemented to other assets due to its flexibility. However, there are some downsides to this model: lack of consideration for other indicators and outside factors. Due to the limited time, not all market indicators are analyzed and tested to generate trade signals based on our price predictions. Considering more indicators could increase the total return from our model. Furthermore, outside factors, like the social media, could significantly affect the short-term price points for the assets. Without such consideration with our model may lead to a huge loss in asset value.

	Market	Baseline	Lin Reg	Poly Reg	ARIMA
Exponential Ratio	**72,403.13**	**106,553.17**	**270,851.13**	**429,818.31**	**609,099.83**

Table 2: Uppermost Results